



# COMPLEXITY CHALLENGES IN ATM

**ComplexWorld.eu**





# Contents

Executive summary.....	<b>4</b>
1 Introduction .....	<b>5</b>
1.1 Purpose of the document .....	<b>5</b>
1.2 Intended readership.....	<b>5</b>
1.3 Inputs from other projects.....	<b>5</b>
1.4 Glossary of terms and acronyms .....	<b>6</b>
2 Objectives and methodology .....	<b>7</b>
3 Expert panel .....	<b>8</b>
4 Documents reviewed .....	<b>9</b>
5 Experts assessment on complexity challenges .....	<b>12</b>



# Executive Summary

After more than 4 years of activity, the ComplexWorld Network, together with the projects and PhDs covered under the SESAR long-term research umbrella, have developed sound research material contributing to progress beyond the state of the art in fields such as resilience, uncertainty, multi-agent systems, metrics and data science.

The achievements made by the ComplexWorld stakeholders have also led to the identification of new challenges that need to be addressed in the future:

- A.** Developing and demonstrating new metrics in ATM
- B.** Building resilience into systems design taking into account emergent behaviour
- C.** Understanding trade-offs through metrics
- D.** Data science and managing and visualising (big) data

- E.** Integrating multi-agent systems into decision-support tools
- F.** Integrating uncertainty into decision-support tools
- G.** Characterisation of meteorological uncertainty
- H.** Model-based identification of emergent behaviours at the design stage, including comparison with reality

In order to pave the way for complexity science research in Air Traffic Management (ATM) in the coming years, ComplexWorld requested external assessments on how the challenges have been covered and where there are existing gaps. For that purpose, ComplexWorld, with the support of EUROCONTROL, established an expert panel to review selected documentation developed by the network and provide their assessment on their topic of expertise.

# Introduction

## 1.1 Purpose of the document

The objective of this report is to present the experts panel assessment on the following complexity challenges in ATM identified by the network:

- A.** Developing and demonstrating new metrics in ATM
- B.** Building resilience into systems design taking into account emergent behaviour
- C.** Understanding trade-offs through metrics
- D.** Data science and managing and visualising (big) data
- E.** Integrating multi agent systems into decision-support tools
- F.** Integrating uncertainty into decision-support tools
- G.** Characterisation of meteorological uncertainty
- H.** Model-based identification of emergent behaviours at the design stage, including comparison with reality

## 1.2 Intended readership

The target reader is any stakeholder interested in complexity research applied to air transport.

## 1.3 Inputs from other projects

The following projects and PhDs have been reviewed with the objective defined above:

PROJECTS	PHD STUDENT'S THESIS
ASHICS	Bouarfa, Soufiane
CASSIOPEIA	Fleurquin, Pablo
COMPASS	Heidt, Andreas
ELSA	Monechi, Bernardo
EMERGIA	Sauer, Manuela
IMET	Schwithal, Alexander
MAREA	
NEWO	
ONBOARD	
POEM	
ROBUST ATM	
SecureDataCloud	
TREE	

Public information on these projects and PhD thesis can be found on the ComplexWorld Wiki (<http://complexworld.eu/wiki/Main>) and/or the SESAR website (<http://www.sesarju.eu/exploratory-research>).

## 1.4 Glossary of terms and acronyms

<b>ABM:</b>	Agent Based Modelling
<b>ABMS:</b>	Agent Based Modelling and Simulation
<b>A-CDM:</b>	Airport Collaborative Decision Making
<b>ANSP:</b>	Air Navigation Service Provider
<b>AO:</b>	Aircraft Operator
<b>ATFM:</b>	Air Traffic Flow Management
<b>ATM:</b>	Air Traffic Management
<b>CDM:</b>	Collaborative Decision Making
<b>CW:</b>	ComplexWorld
<b>DDR:</b>	Demand Data Repository
<b>DST:</b>	Decision-Support Tools
<b>CODA:</b>	Central Office for Delay Analysis
<b>EC:</b>	European Commission
<b>ECAC:</b>	European Civil Aviation Conference
<b>ETA:</b>	Estimated Time of Arrival
<b>EWf:</b>	Ensemble Weather Forecast
<b>FAB:</b>	Functional Airspace Blocks
<b>FABEC:</b>	Functional Airspace Blocks Europe Central
<b>FCFS:</b>	First come, first serve
<b>FMS:</b>	Flight Management System
<b>GPU:</b>	Graphics Processing Unit
<b>KPA:</b>	Key Performance Area
<b>KPI:</b>	Key Performance Indicator
<b>LTER:</b>	Long Term and Exploratory Research
<b>MAS:</b>	Multi Agent Simulation
<b>MET:</b>	Meteorological
<b>NM:</b>	Network Manager
<b>NWP:</b>	Numerical Weather Protection
<b>SVM:</b>	Support Vector Machine
<b>TAAM:</b>	Total Airspace and Airport Modeller
<b>TMA:</b>	Terminal Manoeuvring Area
<b>UDPP:</b>	User Driven Prioritisation Process
<b>US:</b>	United States
<b>WP-E:</b>	Work Package E

# Objectives & methodology

The ComplexWorld consortium agreed a series of research themes which were considered to be important in the context of applying complexity science to ATM, but which still included open areas of research – i.e. that presented research challenges to the community of scientists engaged in this field. Furthermore, a number of open, pertinent research questions were devised by the ComplexWorld partners for each research theme. This list of 8 complexity challenges and the research questions proposed for each on them was therefore identified by the ComplexWorld partners.

SESAR long-term research projects that had addressed these challenges were then chosen for review based on the assessment of a EUROCONTROL project officers' panel. Each selected project was then informed about the research theme and corresponding questions, and invited to submit two works of research (including, at least, one paper) that had developed this theme and explored such questions. It is to be noted that where projects were not selected for inclusion, this does not indicate shortcomings of such projects – the focus was on alignment with, and exploration of, the selected themes and questions.

An independent panel of experts, with domain expertise in the corresponding themes, was then appointed by the ComplexWorld team to comment on the extent to which the selected projects

had addressed the chosen themes and questions. Some panel members were ATM domain experts, others were purposively chosen to bring in views from outside the ATM domain. The panel members were also at liberty to refer to other projects of which they were aware, and to add their personal expertise to their reports. It was not possible to render each panel representative of academia or industry stakeholders. But, for example, a panel could be represented more by ANSP experts than from other domains. The ComplexWorld team limited its intervention to integrate all the panelists' reports into a consolidated report for each theme, further homogenising the reporting styles and seeking clarifications where needed. The consolidated reports were then reviewed and approved by the original panelists.

This exercise was not an assessment of the projects and nor did the resources allow for a fully comprehensive or systematic review. Rather, the objective was clearly focused on identifying particular research challenges going forward and ways in which the state of the art might be improved by building on the existing bodies of work and deploying new ideas and methods.

For the sake of neutrality, this report does not include a conclusions section. Therefore, all opinions and statements on achieved results and future research avenues remain entirely from the external experts.

Bazzani, Armando, University of Bologna

Bhattacharya, Raktim, Texas A&M University

Brenguier, Jean-Louis , Météo-France

Buchanan, Piers, UK Met Office

Havlin, Shlomo, Bar-Ilan University

Lamanna, Fabio, Freelance Civil Transportation Engineer

Lay, Elizabeth, Calpine

Mylne, Ken, UK MetOffice

Netjasov, Fedja, University of Belgrade

Nikolic, Igor, Delft University of Technology

Plets, Hans, Belgocontrol

Reichmuth, Johannes, DLR

Rodríguez Fonollosa, José Adrián, Universitat Politècnica de Catalunya

Servedio, Vito D.P., Sapienza University of Rome

Sharpanskykh, Alexei, Delft University of Technology, the Netherlands

Titely, Helen, UK MetOffice

Vilaplana, Miguel, Boeing Research and Technology Europe

Watson, Mark, UK NATS



# Documents reviewed

- 1.** Arranz A., Etxebarria I., Regidor C., Escribano B., 2013. Intelligent Modelling of the Air Transport Network. 3rd SESAR Innovation Days, KTH, Stockholm, 26-28 November.
- 2.** Blom, H.A.P., Bakker, G.J., Can ground-based separation accommodate very high en route traffic demand as well as advanced self-separation? Proc. AIAA Conf. ATIO 2015, Dallas, 22-26 June 2015, AIAA 2015-3180, pp. 1-15.
- 3.** Blom, H.A.P., Bakker, G.J., In search of positive emergent behaviour in trajectory based operations, Proc. 3rd SESAR Innovation Days, KTH, Stockholm, 26-28 November 2013, 8 pages.
- 4.** Blom, H.A.P., Stroeve, S.H., Bosse, T., Modelling of potential hazards in agent-based safety risk analysis, 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM2013), June 10-13, 2013, Chicago, Illinois, USA.
- 5.** Campanelli B., Fleurquin P., Eguíluz V. M., Ramasco J. J., Arranz A., Etxebarria I., Ciruelos C., 2014. Modeling Reactionary Delays in the European Air Transport Network. 4th SESAR Innovation Days, UPM, Madrid, 25-27 November
- 6.** Cheung J., Brenguier J-L., Heijstek J., Marsman A., Wells H., 2014. Sensitivity of Flight Durations to Uncertainties in Numerical Weather Prediction. 4th SESAR Innovation Days, UPM, Madrid, 25-27 November
- 7.** Cheung J., Hally A., Heijstek J., Marsman A., and Brenguier J-L, 2015. Recommendations on trajectory selection in flight planning based on weather uncertainty. 5th SESAR Innovation Days, UNIBO, Bologna, 1-3 December
- 8.** Clare G., Richards A., 2012. Air Traffic Flow Management Under Uncertainty: Application of Chance Constraints. ATTACS 2012.
- 9.** Clegg, K., Alexander, R., Method description Technical Report, ASHICS deliverable 2.2, University of York, September 2012.
- 10.** Clegg, K., Alexander, R., Model Analysis Technical Report, ASHICS deliverable 3.2, University of York, May 2013.
- 11.** Cook A., Tanner G., Cristóbal S., Zanin M., 2013. New perspectives for air transport performance. 3rd SESAR Innovation Days, KTH, Stockholm, 26-28 November.
- 12.** E.02.04 - ONBOARD - D3.1-Algorithmic Framework Document, internal document E.02.04, 2011
- 13.** E.02.04-ONBOARD-D2.1-State of the Art Review, internal document E.02.04, 2011
- 14.** E.02.06-POEM-D6.2-Final Technical Report, internal document E.02.06, 2011
- 15.** E.02.06-POEM-D6.2-Final Technical Report, internal document E.02.06, 2011
- 16.** E.02.14 - CASSIOPEIA Close-out meeting presentation, internal document E.02.14, 2011
- 17.** E.02.14 - D 4.2 - CASSIOPEIA - Study Report: Case Study 2, internal document E.02.14, 2011

- 18.** E.02.14 - DCI-4HD2D Platform extension report V1, internal document E.02.14
- 19.** E.02.18 – ELSA D2.4 – SESAR Agent Based Model, internal document E.02.18
- 20.** Fleurquin P., Ramasco J.J, Eguíluz V.M., 2013. Systemic delay propagation in the US airport network, *Nature Scientific Reports* 3, 1159 (2013)
- 21.** Fleurquin, P., 2015. Analysis of air transportation using complex networks PhD thesis.
- 22.** Fürstenau N., Heidt A., Kapolke M., Liers F., Mittendorf M., Weiß C. Pre-Tactical Planning of Runway Utilization Under Uncertainty: Optimization and Validation. 5th SESAR Innovation Days, UNIBO, Bologna, 1-3 December
- 23.** Gurtner, G., Lillo, F., Mantegna, R., Miccichè, S., Bongiorno, C., Pozzi S., Adaptive air traffic network: statistical regularities in air traffic management, 11th USA-Europe ATM R&D Seminar, Lisbon, Portugal, June 23-26, 2015
- 24.** Gurtner G, Vitali S, Cipolla M, Lillo F, Mantegna RN, Miccichè S, et al. (2014) Multi-Scale Analysis of the European Airspace Using Network Community Detection. *PLoS ONE* 9(5): e94414. doi:10.1371/journal.pone.0094414
- 25.** Gurtner G., Valori L. and Lillo F., 2015. Competitive allocation of resources on a network: an agent-based model of air companies competing for the best routes. *Journal of Statistical Mechanics: Theory and Experiment*, Volume 2015, May 2015
- 26.** HALA PhD 4th technical report for Alexander Schwithal. 2013.
- 27.** Heidt A., Helmke H., Liers F., Martin A., 2014. Robust Runway Scheduling Using a Time-indexed Model. Fourth SESAR Innovation Days, Madrid, Spain, 25-27 November
- 28.** Heidt A., Kapolke M., Liers F., Fürstenau N., Helmke H, 2014. Pre-Tactical Time Window Assignment: Runway Utilization and the Impact of Uncertainties. 4th SESAR Innovation Days, UPM, Madrid, 25-27 November
- 29.** Kapolke M., Fürstenau N., Heidt A., Liers F., Mittendorf M., Weiß C. Pre-Tactical Optimization of Runway Utilization Under Uncertainty. Submitted to *JATM*, 2015 (NOT PUBLIC)
- 30.** Kapolke M., Fürstenau N., Heidt A., Liers F., Mittendorf M., Weiß C., 2015 Pre-Tactical Optimization of Runway Utilization Under Uncertainty. *Journal of Air Transport Management*. Elsevier. ISSN 0969-6997.
- 31.** MAREA Publishable Summary
- 32.** Monechi, B., Servedio, V. D. P., & Loreto, V. (2015). Congestion Transition in Air Traffic Networks. *PLoS ONE*, 10(5), e0125546.
- 33.** Monechi, B., Servedio, V. D. P., & Loreto, V., 2014. An Air Traffic Control Model Based Local Optimization over the Airways Network, 4th SESAR Innovation Days, UPM, Madrid, 25-27 November
- 34.** NEWO Indicators and Metrics (NOT PUBLIC)
- 35.** Sanchez, M., NEWO Final report, NEWO deliverable D11, Isdefe, September 2013.
- 36.** Sauer M., Hauf T., Forster C., 2014. Uncertainty Analysis of Thunderstorm Nowcasts for Utilization in Aircraft Routing. 4th SESAR Innovation Days, UPM, Madrid, 25-27 November
- 37.** Schwithal A., Hecker P., 2014. Uncertainty Assessment for ETA prediction towards 4D Trajectory Operation. *ICRAT* 2014.
- 38.** SecureDataCloud publishable summary

- 39.** Stroeve S. H., Bosse T., Blom H. A. P., Curran R. Agent-Based Modeling and Simulation of Coordination by Airline Operations Control, to be published in IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING
- 40.** Stroeve S. H., Bosse T., Blom H. A. P., Sharpanskykh A., Everdij M. H. C., 2013. Agent-based modelling for analysis of resilience in ATM. 3rd SESAR Innovation Days, Stockholm, Sweden, 26th – 28th November
- 41.** TREE Metrics (NOT PUBLIC)
- 42.** Zanin, M. (2013). Synchronization Likelihood in aircraft trajectories. In Proceedings of the Tenth USA/Europe Air Traffic Management Research and Development Seminar, Chicago, USA.
- 43.** Zanin, M. (2014). Network analysis reveals patterns behind air safety events. Physica A: Statistical Mechanics and its Applications, 401, 201-206.
- 44.** Zanin, M. et al., Towards a Secure Trading of Aviation CO2 Allowance. Paper submitted to Journal of Air Traffic Management.

**This section comprises the consolidated reports (prepared by the CW team from the experts' assessment) for the challenges enumerated in Section 1.**

## 5.A Developing and demonstrating new metrics in ATM

Within European ATM, there is an opportunity to build on existing work developing and demonstrating new metrics. It is suggested that these should extend the range of flight-centric metrics (e.g. average departure delay) currently used by industry, and cover such performance aspects as cost, resilience, and passenger service delivery. The use of non-classical (including complexity) metrics is expected to continue to play an important role in many instances, although not necessarily required in all cases. Consideration of the complex socio-technical nature of the air transportation system remains underexploited. Improved pathways towards industry adoption of appropriate new metrics are also important.

### **i) To what extent has existing research extended the scope of flight-centric metrics and what are the key new metric(s) areas in need of further development regarding ATM performance assessment?**

Existing research in SESAR long-term research has addressed several of the eleven ICAO Performance KPAs in some

detail. These include, in particular: safety, capacity, efficiency, and predictability – often with useful and inventive project-specific metrics modelled within these KPAs. (Others are less-well addressed overall.)

Such examples include metrics: reflecting the ATM system's ability to control deviations from RBTs and to share this information amongst all users, plus ATM global efficiency improvement measures linked to better information sharing (NEWO); airport delay clusters, whereby the size of the largest connected cluster is used as a measure of the level of network-wide congestion (TREE); and small-worldness and component/cluster metrics used to measure delay propagation effects (POEM). Both of these latter projects deployed full suites of dedicated metrics for reactionary delays.

In ELSA, network community detection algorithms have been used in airspace design, and two metrics have been used specifically to compare new airspace partitions generated by such algorithms:

the Rand index and mutual information. An additional metric, the 'fork', is used in ELSA to express how much a given navigation point is the source of deviation between the planned flight and the actual flown trajectory. Not detectable by conventional metrics, this has suggested that deviations issued/requested are quite repetitive, which has clear strategic (planning) implications.

In addition to classical metrics (such as flight departure and arrival delay) there is an increasing focus on the need to understand delay distributions. In general, there is a current focus on flight-centric metrics, whereas other customer-centric metrics are largely missing. These missing metrics include explicit passenger-centric metrics (POEM being an exception, whereby they are contrasted with flight metrics), and those relating to the freight perspective. Also largely missing are, for example, indicators for: in-delay probabilities for individual flights; door-to-door journey times; connectivity likelihoods as a function of delay duration; fuel-consumption (and emissions) – with detailed sensitivities with regards to weather; staff resourcing and infrastructure capabilities. Cost-effectiveness metrics are also generally lacking at present (see for example EUROCONTROL ATM Cost-Effectiveness (ACE) reports), despite some work on this in POEM. More work also needs to be carried out on cancellations (their costs and impacts), as opposed to delays only. These often improve other performance metrics (e.g. by reducing congestion).

Passenger-centric metrics also need to capture passengers with itineraries cancelled at the point of origin.

**ii) To what extent has existing research been able to demonstrate the value of new metrics? Have empirical (real-world) data been used? What actions are required to move closer towards industry adoption of such metrics?**

Network effects have been analysed (NEWO) by comparing reference scenarios based on FCFS priority rules in use today with alternative prioritisation methods using a mesoscopic simulation tool. The results characterised not only network performance and behaviour, but also the predictability of this performance. Mechanisms for ATFM slot reallocation and swapping (TREE) and various flight-centric and passenger-centric flight prioritisation rules (POEM) have been used to study delay propagation and cost impacts in the European network.

A real day of operations was used as the basis underpinning several of these models, sometimes a longer period. Full, quantitative calibration exercises of the simulation models exist in some cases only, although existing work has often compared given output data (and/or metrics) with real traffic statistics, derived from DDR data and/or sourced from CODA. Measuring absolute impacts on real systems was often not within the intended/possible scope of the existing research.



The tools developed have the potential to be used in real systems as predictors for delays and to measure system impacts, e.g. on airports, flight trajectories and passengers. Comparisons with other network simulation tools would be helpful to demonstrate the strengths and possible limitations of such models.

Research results indicate potential in performance improvements in terms of punctuality, costs and network efficiency. The performance of the ATM system for the passenger in terms of delay and cost can now also be investigated to find solutions that do not degrade system performance as measured in traditional flight-centric metrics. Without the new passenger-centric metrics developed, such optimisation potential would not be visible and therefore never implemented. These opportunities are particularly closely aligned with EU policy developments (e.g. regarding passenger rights).

NEWO considers (localised) unexpected disruptions and the associated impacts on the system. The best network performance occurred with solutions that were good for the airlines: i.e. what was good for airlines was also good for the network. It would be interesting to consider further types of unexpected events in future and how these categories could correspond to compensatory strategies.

Several other questions are prompted regarding further research. Would it be possible to take resilience engineering / complex systems concepts, such as saturation, and also translate these into metrics? Could we measure how often the condition of saturation is reached on certain routes, for example studying groups of real disruptions to determine general shapes of risk and perturbations, patterns and rates of change, to determine what we could learn from them? A useful measure of resilience could look at the

number of times plans change or levels of defences that are moved through, and how these are linked to (internal and external) review processes.

Non-experts in the specific analysis methods used in several research initiatives have to be carefully briefed in the use and interpretation of the results in order to improve the acceptance of powerful new metrics. Regular reporting using the new metrics in combination with the traditional ones would help the industry to detect the cost saving potential for various stakeholders, for example.

Applications that would allow a further comparison of such simulation results with reality (e.g. running the simulations in a shadow mode using empirical data) and also with different flight prioritisation schemes in parallel (e.g. in 'what-if' mode), would probably also increase the chance for adoption of superior schemes by industry.

### **iii) What new methods and approaches are needed - is it clear that complexity science methods have something to contribute towards improved ATM performance assessment?**

"Complexity science postulates that generative rules and equations can be discovered that are capable of explaining the observed complexity of the "real" world/universe. Furthermore, these laws have the potential to predict and control the behavior of real-world systems [...] For instance, it has its hard-core assumptions (e.g., complex dynamic systems can be modeled with generative rules; similar generative rules operate across a wide range of complex systems) and methods (e.g., agent-based computational modeling; non-linear dynamics; genetic algorithms)."

Traditional methods used to evaluate ATM performance are unable to keep up

with the system's complexity. It is clear that complexity science can contribute to improved ATM performance given that air traffic systems are the epitome of complex systems with many interdependencies.

There are distinct components of research delivery already highlighting the benefits that complexity science brings to improved ATM performance assessment. In NEWO, a simulation tool has some complex generative rules implemented that seem to produce similar behaviour at the network level, as observed in reality. Similar observations have been promisingly made in other SESAR long-term research work. Using TREE's metrics for air transport delay propagation (e.g. the cluster of congested airports indicator) will certainly lead to deeper insight into network delay mechanisms, and may indeed lead to an indicator that supports decision-making for aircraft delays in real-time operations and support airport arguments for capacity extensions. Airline-decision making based on cost rules and passenger/flight performance is a key focus within POEM, with supporting analyses on specific airport roles in the network.

Critical considerations are emergence and interactions across scales: we still need more rules that work at different scales and improved characterisations of delay nodes. Furthermore, resilience is not about state but about how things are moving, such that we need to be sensitive to rates of change - are we working harder to stay in control? A core tenant of resilience engineering is that adding flexibility to a system increases resilience. Rates and shapes of change could be considered further, for example: are there pauses in delay propagation? Where is propagation fastest or slowest? Relationships between system components also need to be explored and how these can be better coordinated (a type of network-wide CDM).

Excellent and detailed discussions can be found in certain outputs regarding the use of traditional analysis methods compared with complexity science methods (such as Granger causality analyses in a directed network). The combination of both approaches knowing their strengths and limitations leads to deeper insights into the mechanisms that steer network performance. The Granger causality analysis detects causally-connected airports in delay propagation in the network, that were not detected by using classical statistical analysis methods. Direct comparisons of results using 'conventional' methods, e.g. microscopic simulation fast-time tools, could well be useful, as also elaborated in the previous section.

SESAR long-term research has also found that giving less weight to network-driven priorities provided better network performance. This is in alignment with resilience engineering observations that polycentric governance (with partial authority, autonomy and high-level goal alignment at multiple centres of adaptive behaviour) is effective in managing complex systems. How much further could such governance be delegated downwards, e.g. in contrast to FAB objectives?

All systems have boundaries and finite resources. System brittleness is defined by the behaviour of a system when it gets near to its boundaries. It appears that another area of ATM interest could be behaviour near its operational boundaries.

**iv) What facets of ATM and wider air transport system assessment are currently excluded from performance models? For example, is more work required to embrace the complex socio-technical aspects of the system?**

Performance assessment could be extended specifically to better cover airline and airport stakeholders, including freight

operations, and also societal viewpoints. Turnaround and terminal processes at airports are not yet covered in depth, either. Flight prioritisation rules need further discussion regarding impacts on equity and user participation, in addition to global interoperability and the legal aspects of oversight for correct application and the control of misuse. Also, the integration of explicit flight trajectory models into current planning tools may help to extend the application of some existing research models more in the direction of 'what-if' tools for the network manager.

Consideration of possible flight/passenger demand changes both in terms of location and time in the future could give insights into impacts on ATM performance in the ECAC area, e.g. if the global network changed its centre of gravity more towards Asia.

Generic organisational and cost models (including profitability) of different stakeholder types may be useful to obtain a reference standard and wireframe model to allow the assessment of future developments. Extending metrics (e.g. to embrace passengers), providing stakeholder-specific sets of indicators, can be developed in general to determine not only trade-off effects between different interest groups but also to search for system configurations that could improve the performance of the transport system in general, even outside the specific ATM domain. There is always more than one perspective and the view from any single point of observation both reveals and obscures. Research needs to shift and contrast across perspectives.

Better understanding of patterns and rates of impact also represent important areas of future research, particularly their dependencies on types of route, carrier and airport. It would be interesting to study couplings (tight or loose), rates

of change, vectors (in what direction does the cascade flow?) and impacts across the system. From there, one could explore how to loosen particularly tight couplings or how to increase the resilience of brittle system conditions. In many cases, reference standards and empirical baseline data (especially from airlines) are not yet available.

## 5.B Building resilience into ATM system design taking into account emergent behaviour.

A key challenge is how to make the ATM system more resilient regarding disturbances and disruptions? For the complex socio-technical ATM system this comprises the following three complementary contributors:

- **Robustness** is well known in systems engineering for the **absorptive capacity** of disturbances by a technical system.
- **Engineering resilience** has its focus on the **restorative capacity** of a system in terms of the time required to recover from a disruption.
- **Resilience engineering** has its focus on the **adaptive capacity** of a safety-critical socio-technical system in response to disturbances and disruptions.

Because each of the above mentioned resilience capacities are of great value to ATM, the question is how to build them all three into ATM system design, and how to measure these capacities in an objective way?

This resilience performance question is, however, only one side of the ATM performance medal; the other side consists of established key performance areas such as economy, capacity and safety. Therefore, we are in need of an ATM system design that is more resilient against disturbances and disruptions and at the same time maintains a good balance with other key performance areas.

In support of a step change in future ATM design, this challenge concerns Building resilience into systems design taking

into account emergent behaviour. This evaluation report aims to describe for this challenge, which relevant achievements have been made within relevant SESAR research [1-6], and what the follow-up research directions are.

### i) How well has SESAR funded research addressed the topic of how to measure/quantify resilience of the ATM system?

Within SESAR funded research, a good start has been made in taking a complexity science perspective to the study of the three resilience capacities in the ATM system. Regarding robustness, in [1-2] it is shown how such a metric can be used to capture, to optimize and to measure the absorptive capacity by ATM of uncertainties in aircraft flight arrivals and departures at a busy node in the air transport network. In [3] a literature review of various domains has been conducted regarding the availability of resilience views and metrics that go beyond those of well-known system engineering. The key finding is that a large variety of such resilience metrics has been developed, though that their relation with established KPAs is unclear. It was also identified that none of these metrics is able to capture the effect of the adaptive capacity of a socio-technical system separately from the capture of the effects of the absorptive and restorative capacities. In [4] it has been shown that an effective way to address this problem is to develop a proper agent-based model of the socio-technical system considered, and subsequently measure the performance using two types of simulations: one for the full model, and the other for a version of the model in which the adaptive

capacity is nullified. In [5-6] the focus was on agent-based modelling various types of hazards at cognitive, individual, social, and organisational levels that can be used to capture resilience of a socio-technical ATM design. This is promising because in modern ATM systems resilience is often achieved by interaction between social (e.g., pilots, air traffic controllers) and technical system components. In particular, the adaptive capacity of resilience relies heavily on human contribution in modern ATM systems. Therefore [5-6] may be expected paving the way to the development of novel mechanisms to improve resilience of ATM systems.

Relevant further developments regarding characterizing different patterns of disruptions, using real data as well as model simulation studies, could be addressed by future research, specifically in the following two main directions:

- The identification and quantification of resilience metrics at different aggregation levels (individual, team/unit, organisational, inter-organisational) and establishing relations between these metrics across different levels. In such a way the emergence of resilience in a global ATM system could be investigated and resilience could be monitored at different aggregation levels.
- Continuing the development and implementation of simulation models of complex ATM designs that have the potential of revealing and analyzing inter-dependencies in the complex system which are highly relevant for identifying ATM system vulnerabilities for different types of disturbances, and addressing resilience performance as well as other key performance areas.

**ii) How well has SESAR funded research addressed the topic of how to make the ATM system more resilient and what should be done here in the future?**

The issue of making the future ATM system more resilient has been addressed in a few SESAR funded studies. In [1] and [2] mathematical approaches have been studied regarding the joint optimization of robustness and efficiency through improving the absorptive capacity of uncertainties in flight arrivals and departures at a node in the air transportation network. More specifically, this activity proposed to use stochastic and robust optimisation techniques to find robust solutions taking into account uncertainty in runway utilisation. In [4], coordination mechanisms have been studied and compared regarding their engineering resilience performance for use in airline disruption management. The activity delivered a novel coordination approach based on a theory from social sciences, which resulted in a fast, cost-efficient recovery by an airline from a significant disruption.

The robustness and engineering resilience types of results obtained in [1,2,4] show that formal and computational models are of great value in understanding, predicting and improving complex nonlinear dynamics of ATM systems with their numerous interdependencies, threshold effects, and feedback loops operating at multiple temporal and spatial scales. This forms a good motivation to strongly continue these mathematical optimization and agent-based coordination approaches in future research. The key issue is to handle various uncertainties and changes in the ATM system which combine elements of hierarchical top-down control and bottom-up self-organisation. On the one hand, if a system is tightly connected and strictly controlled, it may be too rigid to handle disruptions in a timely and efficient manner. On the other hand, if it is too loosely connected it may decouple in weakly connected parts, which would function inefficiently at a global, systemic level. In future research it may be explored



how different mixes of centralisation and autonomy of human and technical system components of ATM systems influence the resilience of the whole system in different types of environmental dynamics. To enable such analysis formal mathematical and computational simulation tools need to be further developed and used. In doing so it is important to consider interdependencies among its different actors: airlines, airports and air navigation service providers. In the future, models and methods addressing these interdependencies at different aggregation levels need to be developed. Also important is to focus on ways of ensuring the representativeness of the set of scenarios under consideration with respect to the real world dynamics, but at the same time ensuring computational feasibility of optimisation. Different sources of uncertainty need to be identified, quantified and justified by real data. The issue of scaling up of approaches to ensure robustness of large ATM systems (e.g., in case of a major disruption) needs to be addressed in the future too.

In future complexity science research, mechanisms of the adaptive capacity also require further exploration. In particular so because the adaptive capacity of resilience of complex sociotechnical systems, such as ATM systems, is not well understood. Key areas for development are:

- individual, social and organisational learning;
- anticipation and early recognition of disruptions;
- decision-making under uncertainties in sociotechnical teams;
- adapting an organisation to novel types of disruptions.

Most of the existing studies on such adaptive capacities are descriptive and are based on real cases. Mathematical and computational

models, as well as computer simulation studies are rare. However, such formal tools are indispensable for studying dynamics of complex adaptive systems, such as modern ATM systems. Such models and tools should nevertheless be based on strong empirical evidences of resilience of real high risk organisations, which are currently available in abundance.

**iii) How well has SESAR funded research addressed the topic how to timely identify positive and negative emergent behaviours of future ATM designs at various frequencies, i.e. both the frequent nominal as well as the rare non-nominal behaviours?**

[1,2] have focused on negative emergent behaviour in the form of flight delays due to uncertainties in flight progress. Some of negative emergent behaviours were addressed at the pre-planning phase by incorporating uncertainty in nominal models of operations. Some other negative emergent behaviours were addressed by recovery to strict robust solution technique, which allows repairing some infeasible solutions. It is indicated in the activity that none of the proposed techniques can prevent infeasible solutions completely. Therefore, in the future, approaches need to be developed, which would address resolution of conflicts and coordination between conflicting actors during the actual execution of operations. Such approaches should take into account propagation of conflicts and resolution actions through the system.

In [4] the focus mostly was on positive emergent behaviours that could be achieved by employing the proposed coordination protocol in a case study at an airline Operations and Control center. It is suggested that in the future, more generic techniques and tools would be developed to explore positive emergent behaviors in ATM systems. Such techniques and tools may provide more insights in how

positive behaviors would emerge from local interactions of social and technical systems. Understanding mechanisms and identifying strong contributors to positive behaviours is essential for building new resilient ATM systems. On the one hand, such techniques and tools could be used for detailed offline analysis of novel ATM designs. On the other hand, they could be used for online monitoring of performance of ATM systems and timely identification of emergent properties. In particular, tools for anticipation of critical transitions in complex ATM systems could be developed.

In [5-6] the focus of the activity was on modeling and identification of safety hazards. In follow-up research, these modeling constructs may be used to explore both positive and negative emergent behaviours in agent-based models of future ATM designs, and relating these behaviours to resilience across different aggregation levels in the complex ATM system.

Under challenge H, entitled “Model-based identification of emergent behaviour at the design stage, including comparison with reality”, the very same question regarding emergent behaviour has been addressed for eight other SESAR funded research papers and reports. This provides significant further insight regarding the progress made by SESAR funded research on the issue of emergent behaviour of future ATM designs.

#### **iv) How well has SESAR funded research addressed the topic of how to get a desired trade-off between resilience and other KPAs?**

[1,2] considered trade-off between robustness and KPAs related to delays and the amount of aircraft movement during the recovery phase. Safety was considered by ensuring separation regulations between aircraft of different types.

[4] considered a case study where resilience directed performance was balanced against

other KPAs, such as economy.

In future research, the KPAs reflecting different dimensions of ATM performance (efficiency, safety and others) needs to be balanced with resilience performance metrics. As ATM is a complex, multi-level sociotechnical system, agent-based modelling and simulation is one of the most promising approaches in realizing this. The latter is strengthened by the agent-based models developed in [5-6] of the various hazards in ATM.

#### **References**

1. Andreas Heidt, Hartmut Helmke, Frauke Liers, Alexander Martin, Robust Runway Scheduling Using a Time-indexed Model, Fourth SESAR Innovation Days, Madrid, Spain, December 2014.
2. Manu Kapolke, Norbert Fürstenau, Andreas Heidt, Frauke Liers, Monika Mittendorf, Christian Weiß, Pre-Tactical Optimization of Runway Utilization Under Uncertainty, Draft of Accepted Article for Special Issue of the Journal of Air Transport Management.
3. Soufiane Bouarfa, Henk A.P. Blom, Resilience, Chapter 5 in: Complexity science in air traffic management, Eds: A. Cook and D. Rivas, Ashgate Publishing, ISBN 978-1-4724-6037-0, to appear in 2016.
4. Soufiane Bouarfa, Henk A. P. Blom, Ricky Curran, Agent-Based Modeling and Simulation of Coordination by Airline Operations Control, accepted by IEEE Transactions on Emerging Topics in Computing, to appear in 2016.
5. Sybert H. Stroeve, Tibor Bosse, Henk A.P. Blom, Alexei Sharpanskykh, Mariken H.C. Everdij, Agent-based modelling for analysis of resilience in ATM, Third SESAR Innovation Days, Stockholm, Sweden, 26th –28th November 2013.
6. Publishable Summary of the MAREA project, NLR, 2013.

## 5.C Understanding trade-offs through metrics

Current (Key) Performance Indicators and the trade-offs between them are not sufficiently understood, especially in terms of stakeholder impacts, such as costs. Some established work has been carried out by EUROCONTROL on the trade-offs between en-route capacity provision and ATFM delay, but this represents one of few such examples. Trade-offs between monetised and non-monetised metrics are particularly challenging.

**i) To what extent are the horizontal trade-offs (i.e. within any given period) understood, between stakeholders (e.g. ANSPs, airlines, airports, passengers and airlines) and also regarding environmental impacts? What further research is required?**

In general, horizontal trade-offs are not well quantified in most current research because KPIs that truly reflect the goals of each stakeholder are not all available. Some trade-offs need a consensus across stakeholders; e.g. regarding the importance of predictability in ATM operations c.f. other KPIs (excluding safety). There is a direct conflict between the goals of some stakeholders, both within and between such stakeholders. ATM wants predictable, planned operations in order to minimise the complexity-per-flight for controllers, for example, while AOs want to take all possible tactical opportunities to reduce fuel burn. The operator of an airport wants high throughput in typical conditions, minimal arrival delay, resilience in adverse conditions, and predictability. However, maximising throughput probably entails the acceptance of some TMA holding, and makes it harder to achieve resilience and predictability. Airport operators themselves

do not well understand the trade-offs between these goals. Consider further the apparent win-win of environmental noise reduction with steeper glideslopes. Here, there may be a penalty on the airframe (harder landings) which cost airlines more in maintenance and it may not be so easy to keep a consistent stream of arrivals and hence capacity may be adversely impacted. Therefore in this example, the airport gains environmentally at the penalty of capacity and the airline partly loses financially, too. Even with A-CDM at an airport, the perennial trade-off of local versus network optimality arises.

SESAR's long-term research has studied a number of trade-off examples. Quantitative trade-offs are made between the environment (noise) and cost efficiency (economic impact) KPAs in CASSIOPEIA through studying the effects of imposed curfews in terms of economic impacts on airlines, airports and the local (Frankfurt) economy. CASSIOPEIA's second case study examined a collaborative decision-making mechanism in which en-route regulated slots could be swapped between airline operators through bidding and slot selling. Trade-offs are thus established between delay (KPA: capacity) and cost (KPA: cost efficiency), as in a further case study (using agent-based modelling) to understand the impact of an increased use of variable aircraft speeds into a major European hub. Environmental impact indicators (CO<sub>2</sub> emitted – total and per flight) were also taken into account, which is uncommon in such research.

Inter alia, ELSA studied deviations between planned flights and actual trajectories. The lower the deviations, the higher the

predictability and punctuality, while higher deviations – typically an actual flight being allowed to flow more directly than planned, thus omitting some navigation points – usually translates into lower fuel burn (with a positive impact on cost and emissions). In addition, the structure of the navigation point network impacts the complexity and hence the controller's workload.

Airlines tend to react to a downturn in passenger demand by reducing the frequency of service: such strategies are commonplace and tend to require compensatory higher load factors. Where AOs have taken such action this then impacts seat capacity reserves, and this is where the current relationship between the passenger and AO stakeholder trade-offs can be more clearly observed. From a passenger-centric standpoint, seat reserves are one of the principal factors that impact passenger delay. A better understanding of the trade-offs between passenger-centric and flight-centric metrics is presented by the POEM project, which also includes cost impacts. It has here been demonstrated that certain flight prioritisation rules may save airlines money through improved service delivery to the passenger, even though these improvements may be statistically invisible to flight-centric metrics such as average arrival delay. The impact of flight prioritisation rules on ANSPs (e.g. controller workload) would be interesting to model (further) in future, and to understand the AO equity impacts of priority routing / sequencing for certain flights (e.g. with high passenger connectivities).

**ii) To what extent are the vertical trade-offs between periods understood, particularly looking ahead, well beyond 2020? What further research is required?**

There is a strong body of research now published that indicates the importance of stakeholder engagement and a need for them to make clear their long-term and

short-term business drivers, objectives and goals in the ATM system. The vertical trade-offs are understood in principle, but are difficult to quantify because of the uncertainties in costs (development, implementation/equipage, and operating) and benefits. The case for using new airborne capabilities in ATM is not clear, and this is a particularly important consideration because of the lead times involved in the widespread introduction of new FMS capabilities.

Uncertainties extend more generally to future aircraft fleets (including their environmental impact and fuel consumption) and airports - over longer timescales, an airport may change its number of runways, its layout or, in some cases, new/alternative airports may open. Consideration of possible flight/passenger demand changes both in terms of location and time in the future is also necessary and could give insights into impacts on ATM performance in the ECAC area, e.g. if the global network changed its centre of gravity more towards Asia.

Research is needed not to create a single stakeholder arrangement with one particular timeframe in mind but rather to create tools that can be used to represent future market trends, future policies in ATM and to investigate the impact of a multitude of potential scenarios. Relatively little quantitative work has been done in this direction so far.

**iii) What discrepancies are identified between regulatory (e.g. SES Performance Scheme targets) and business (market) forces?**

Regulatory impacts have been assessed in SESAR's long-term research to a fairly limited extent, overall. Some examples include: airport curfew restrictions and their impacts on the airlines' business; the impact of Regulation (EC) No 261/2004 on passenger disruption recovery rules;

and, ATFM slots traded as a commodity under postulated, regulated auction frameworks and their impact on airline cost performance. In the latter case, it has been demonstrated (e.g. by CASSIOPEIA) that, under certain conditions, there are mechanisms that can be beneficial in terms of business impacts (e.g. delay reduction).

Using community detection algorithms, ELSA could be seen as an (unsupervised partitioning / free design) tool to shape the regulatory environment in a way to better match operational requirements, and hence business forces. The obtained partitions provide alternative (bottom-up, apolitical) ways of designing European airspace, with cross-national control units based more on traffic demand than on national constraints. This has the potential to propose, for example, new ACCs, which would be more densely connected inside and have less interface with the exterior, which is good from an operational point of view and reflects the philosophy underpinning the Single European Sky initiative. On the other hand, it has also been demonstrated that the design of the currently existing Functional Airspace Blocks (FABs) might not be optimal (e.g. regarding FABEC there is little operational basis to have France and Germany in the same FAB, it has been demonstrated).

The POEM model focused on analysing the discrepancy between the current flight-centric approach and the desired, more balanced approach, in which the importance of the passenger as stakeholder in the aviation value chain is recognised through the introduction of passenger-centric metrics. A discernible conflict between regulatory and business forces could become more apparent as soon as a (sufficiently strong) revision of Regulation 261 is implemented. This model would also enable and facilitate the incorporation of passenger-centric metrics into the SES performance scheme for RP3. There is,

furthermore, a well-defined place for flight prioritisation strategies within UDPP (user-driven prioritisation process); this is the ideal environment for passenger-focussed prioritisation processes to be investigated.

Conflicts requiring more research to understand the corresponding trade-offs, are as follows:

- The regulatory requirement (formally on Member States, but practically on ANSPs) to achieve performance targets, especially in the Key Performance Areas of capacity and environment, can be in conflict with AOs business forces to reduce cost. Indeed, while ANSPs provide shorter routes to improve horizontal flight efficiency, AOs can decide to fly longer routes, as their cost model optimises their costs, taking into account not only nautical miles flown, but also factors such as unit rates and weather. In other words, the shortest route is not necessarily the cheapest route.
- The regulatory requirements of the SES Performance Scheme can drive an ANSP with lower delays than targeted to decrease the manning of its sectors, hence optimising its own cost, while generating more delays for the AOs. The regulatory requirements can hence lead to a conflict between an ANSP's and an AO's business objectives.
- ANSPs invest money in improvement projects which generate benefits for the AOs. The implementation of an improvement project may lead to lower revenues for some of the ANSPs concerned as it creates new routes which shift traffic geographically. There is hence a conflict between an ANSP's business driver to control costs and their regulatory requirements to improve operational and safety performance that serve the AOs.
- SES Performance Scheme targets are set



five years in advance, as these periods are of five years' duration (except for the first one). They are hence based on assumptions – especially regarding traffic evolution – which can substantially deviate from reality.

- With extended AMAN, the tasks of a downstream ANSP become more complicated in order to assist a distant airport; similarly effects apply to ground holding at the origin airport, in order to respect target times along the route, which may thus disrupt stand planning.
- In addition to metrics for overall system performance, metrics that reflect the interests of each stakeholder are needed in order to predict the effects of new procedures and to determine where Implementing Rules are needed.

**iv) Which types of metrics are missing from such trade-off analyses? - Measures of resilience? Non-cost metrics, such as delay variability and NOx emissions?**

Within the reviews for this particular challenge, the projects were not focused on safety assessments. Whilst it is recognised that quality work on safety assessment has been undertaken elsewhere within SESAR long-term research, safety effects are often apparent in other projects (for example with regard to aircraft diverting to less familiar airports, pressures of punctuality, changing speeds at certain parts of the flight path, schedule changes etc.), but not explored. These effects could in future be at least qualitatively assessed, even where resources do not permit a quantitative assessment.

Several existing long-term research models could accommodate emissions estimations, although few currently consider them (CASSIOPEIA being one exception, as cited above). Noise footprints on the ground around airports are rarely addressed

sufficiently, nor are the wider societal impacts of aviation.

Measures of resilience are becoming more common, from a low base, but generally not yet mature or supported by calibrated data. This area is starting to evolve in SESAR long-term research and Horizon 2020.

In addition to classical metrics (such as flight departure and arrival delay) there is an increasing focus on the need to understand delay distributions (see challenge: 'Developing and demonstrating new metrics in ATM').

**v) Have empirical (real-world) data been used? What new methods and approaches are needed (such as data-driven and predictive analytics)?**

Agent-based modelling (simulations) has been variously demonstrated to be an alternative to analytical methods, showing that it can handle more complex strategies and exchange mechanisms, including the representation of multiple stakeholders. The novelty of such approaches is that they offer a new framework for ATM performance modelling.

To achieve realistic outputs, a good understanding of the possible (input) tactics/ influences are required, ensuring that any agent decisions are rational. The strength of the results of such models depends on the initial, detailed understanding of the problem and associated agent behaviours. There is scope for more work in testing the alignment of agent-based rationales with stakeholder intentions, and comparing outputs directly with other (analytical) methods.

Empirical data have been used as inputs (e.g. traffic data (by far the most common, as readily available from EUROCONTROL through PRISME / DDR), fuel prices, airport curfew data, passenger itineraries and connectivities (much less common)).

DDR data have been used for the last-filed flight plan (to best represent the planned trajectory) and radar data (to best represent the actually flown trajectory). Data describing the structure of the airspace (NEVAC files) have also been used.

Modelling approaches have been supported by tools and methods adopted from complex systems, complex network theory, random graph theory, game theory and affine disciplines. Several SESAR long-term research projects treat the ATM system as a multi-layer, multi-scale system, which calls for analysis methods that are provided by network science. Analytical methods of note include:

- (unsupervised) network community detection algorithms, used in airspace design;
- multi-resolution modularity, graph perturbation and comparison of partitions across operational days, used to assess the robustness of new airspace partitions (designs);
- factor analysis (principal component analysis) to describe delay propagation;
- an adaptation of the Kaplan-Meier estimator for the survival function, to compare flight prioritisation processes;
- Granger causality analyses, to detect the presence of causal relationships between time series.

Future SESAR long-term research should, in general, take actually flown routes and user-preferred routes more into account, which would serve as good benchmarks for assessing future performance designs and enabling user-driven solutions. Further consideration in bottom-up modelling could be given to the fact that current/actual traffic flows already reflect the constraints imposed on existing flight planning and airspace use, which themselves result

from a top-down approach. In deriving conclusions from purely data-driven approaches, researchers should avoid overlooking simple operational constraints that explain certain observed patterns in empirical data. (An example is diurnal variations in the number of navigation points being linked to the booking and releasing of military training areas, which will vary from state to state.)

Some modelling (e.g. of departure slots) will need in future to take more explicit account of (local) operational constraints (e.g. pre-departure sequences at A-CDM airports). A key question is how does airport resilience align with various (new) flight prioritisation strategies? Established models (such as specific airport movement and configuration models, e.g. TAAM) would need to be cooperated with newer models to reliably assess these effects.

There is also a need for new models that take controller decision-making and airline planning intent and rationales into consideration. Finally, it is noted that predictive analytics are currently largely underexploited, although some stakeholders (mostly airports) are acting as vanguards in this domain.

## 5.D Data science and managing and visualising (big) data

Data science techniques together with complex systems theory and practice opens a new approach in the study of the complexity of the Air Transport. Significant research challenges are identified in this field, like data management, data processing, data sharing and protection, deep analytics or visualisation. For aviation to access and manage the datasets generated by the different agents, suitable data infrastructure paradigms need to be developed. Extracting knowledge from data that represents, predicts and improves the behaviour of the system, requires collecting, validating, formatting, correcting and consolidating different datasets. Considering the heterogeneity of the data sources (aircraft, airlines, passengers, navigation services, ground handling, retail sub-systems...) the management of big data can be considered a complex challenge in the aviation field. Even more, if we consider the volume, variety and velocity of the datasets. Other techniques barely explored in aviation, like data protection paradigms or data visualization can be enormously helpful in the field of Air Transport, ensuring the analysis of the performance, the use of existing resources and the support to the decision making processes can be improved several orders of magnitude.

### **i) To what extent researchers use a significant volume and variety of data to derive data-driven paradigms?**

Below the umbrella of ComplexWorld projects and PhDs, there are very few examples of projects dealing with a significant volume and variety of real data. Privacy and confidentiality concerns on the sharing of sensitive, private data limits the research in this area. In this line, the use of secure computation techniques

in data science, such as secure multi-party computation as developed by SecureDataCloud SESAR Long-term research project, can be the key to solve this important practical problem in big data. Using confidential information for data analysis and machine learning without revealing raw data or individualized information is very valuable in the aviation field.

Both COMPASS project and P. Fleurquin PhD thesis manage a database with significant volume and variety of data to derive data-driven models and machine learning techniques in the case of COMPASS. For the analysis of safety events, the unbalance is an unavoidable characteristic of the database, therefore, specific algorithms or training strategies need to be chosen.

In order to enhance the variety of data considered, it would be interesting to combine data from different airspaces with different regulations (f. i. European and US databases regarding re-scheduling in the case of delays).

### **ii) Are the current research projects building predictive analytics capable of deriving forecasts to different time horizon to feed any operational paradigms?**

Yes, predictive analytics have been built in the fields of safety and flight delays.

In the field of safety, existing research provides a new way to forecast not only the simple occurrence of an unsafe event, but a level of Probability of occurrence, which is actually a significant improvement in air safety management. The proposed approaches are actually directed straight to help both human and automatic Air Traffic

Controllers. Future research in operational procedures can consider the integration of the developed measures and others in a joint global prediction system of unsafe situation in different time horizons.

In the field of flight delays, the developed flight delay model can provide the complete temporal evolution of clusters of congested airports for each day, given the initial delays and the schedules of that day (initial conditions). The model can also be modulated to study the effect of different initial conditions in network congestion: daily schedule, plane rotation, flight connections of either passengers or crews and airport congestion. The topological analysis of the air transportation system is based on a fixed network. But in future work it can be extended to show and to predict the evolution of the network and the resulting communities.

**iii) Are the current research projects using the latest findings in machine learning and knowledge discovery? Any particular and promising techniques are not being used that could have some potential?**

Machine learning techniques have only been applied for the prediction of unsafe situations. In particular, a Support Vector Machine (SVM) algorithm was selected to classify network using different topological metrics as input features. Although this choice was considered appropriate for the scope of the research, other machine learning algorithms as Tree Ensembles (Random Forests and Gradient Boosted Decision Trees) were suggested for future work. The innovative network approach proposed in this work models aircrafts in the same radius of influence as nodes and connect them. This is one of the first approaches that goes towards an operational view of a network rather than the use of pure topological systems (which usually associate a weight to edges to take into considerations dynamics as travel time or frequency of services). To complement this work, it is

recommended to combine the network features with other standard dynamic features (such as the Synchronization Likelihood) in a global predictive platform using Gradient Boosted Decision Trees or another advanced machine learning algorithm for heterogeneous inputs.

The model used for flight delay spreading calculation is considered appropriate although it is not based on machine learning and knowledge discovery algorithms. It is particularly interesting the study and the comparison among different clustering techniques (INFOMAP, OSLOM and Louvain) to find and apply the best algorithm that can correctly shape and detect clusters. In addition, the application of machine learning and knowledge discovery algorithms to this work is recommended to extend current research getting an accurate individualized estimation of delays.

**iv) Are the current research projects building or operating big data infrastructures?**

Large amount of data is managed, but no big data infrastructure has been built nor operated.

The prediction of unsafe situations uses a total of 10.3 million flights, containing 100,032 potentially unsafe situations, 4,316 of them unsolved. The flight delays analysis uses a dataset of 6,450,129 flights operated in the US in 2010 by 18 carriers and connecting 305 airports, i. e. 78% of the US flights. Although, at this stage big data infrastructure is not needed, the collection of the data in a future real-time application of the project results will require a complex big data infrastructure. Still, in order to provide the best solution for future needs, it is fundamental starting to explore Big-Data Infrastructure and potentials (e.g. Grid Computing).

**v) Are the current research projects leveraging in visualisation techniques to improve the understanding of data**

### or communication of findings? Are the visualisation techniques helping to increasing the speed of analysing big volumes of data?

Colored graphs on geography maps have been generated to visualize the flight delay propagation. This visual support is considered an important part of the project. For instance, clustering patterns and the visualization of the cascading of delays over time provide instant information about the evolution of short and long term forecasting of flight delays. The visualizations generated are considered helpful for the understanding of the potentials of the methodology and their outcomes, even by non-experts. In the future, the real-time visualization of the spreading of the delays can be very useful to propose adequate solutions.

No visualization techniques have been considered in the COMPASS reports but they can be of interest in the analysis of the networks developed. The derived indicators can be used to complement the information currently visualize in ATM monitors.

### Conclusions

- More research is needed in the field of Data Protection. Secure Multiparty Computation technique is not using big data or machine learning techniques, but it clearly shows that the proposed SMC framework is of great interest in machine learning. SMC is the key to solve one of the most important practical problems in big data applications: data confidentiality. SMC opens the possibility to learning from shared data without an explicit access to the raw data. There is still a lot of work to be done to apply SMC in big data. One of the general challenges limiting the applicability of SMC to real-world tasks is the large computation cost required to perform even very simple analysis.
- No specific data processing architectures exist for the analysis of aviation data.
- Safety research would benefit from a holistic deep analytics approach. The existing research of safety data presents promising new techniques that should help the air safety management in controlling, forecast and prevent unsafe situations due to Loss of Separation events. The methodologies chosen are well supported by literature and full suitable for the application to the air transport safety analyses. Some possible improvements suggested propose comparing datasets of different ATMs approaches on safety: for instance, European vs US data. It would be interesting analysing how factors such as flight scheduling, departing queues or rescheduling in case of delays, along with their respective laws, might influence also air safety.

New hardware solutions based on cloud computing or computational efficiency of GPUs can be of great help in this aspect. It is also necessary to develop new SMC strategies to deal with the huge amount of computation required with advanced machine learning techniques. Regarding the latter topics, some insights and overviews of Big-Data processing (such as Grid-Computing) may be necessary to forecast future needs. Even if the actual amount of data and computational time don't need such architectures, it is fundamental to build solid basis on such topics to get "all the things ready" once data could not be handled with the current technologies. One key question is how effectively secure and reliable is the full process: to answer at this question, it is worth to perform a full test running hypothetical attacks to the system to evaluate the vulnerability of the system and its Achilles' heels. The second question is the attitude of the actors in using and trusting the introduced Graphical User Interface without the need of more actions. In order to provide answer to the latter, a new common "working protocol" would be needed.



- Predictive analytics on flight delay shows results to improve ATM performance. The research based on flight delay data show a deep review of the existing approaches on ATM to model, detect and forecast delay propagation within airport network. The model developed shows a great potential and accuracy in helping ATM stakeholders not only to forecast but, as a potential final outcome, to prevent the occurrence of strong cascading delays over a network providing a better understanding of the operations dynamics. Some dynamics of air traffic might be further explored (or merged) among the same methodology applied on different air spaces regulations (i.e. US and Europe), in order to discover, whether exist, universal patterns in the delay formation. Still, on the other side, a measure of the impact of regulation in the dynamics of the delays propagation might be interesting. Finally, it is highlighted the hidden potential of these methods of air traffic forecasting outside the ATM industry.

## 5.E Integrating multi agent systems into decision-support tools

Multi Agent Systems (MAS) in the Air Traffic Management (ATM) can have important advantages for policy makers as they allow evaluation during the design of a novel operation and they allow performing scenario simulations in terms of what-if studies through tuning the relevant parameters of the model.

Moreover, MAS and Agent Based Models (ABM) can be relevant for the investigation of the behavior of the ATM main actors, as they can provide useful insights about the learning mechanisms on which the agents' behavior is based.

These features can be fruitfully exploited by using integrated decision-support tools (DST), based on MAS and ABM, that will help in selecting the best policies and strategies to improve the general efficiency of the ATM system, building on the analysis of historical data.

A relevant example might be given by the Business Trajectory issue in the future SESAR scenario. In this case, ABM-based DSTs might be used to investigate the policies regarding the rules by which the airline will compete for the selection of the business trajectory that best fits their internal needs.

In order to fix the general framework in which the activities related to Challenge E are laying, it is perhaps useful to report a few considerations done by one of the external experts on Agents Based Models. In his words: *"The ABM approach to manage complex systems has the ambition to mimic the cognitive behavior of individuals by an algorithmic implementation of many different strategies and decision*

*mechanisms. Therefore, the ABMs try to reproduce in silico a complex virtual reality taking into account many different aspects of the information processing, learning capacity and rational behavior of human beings. In such a way, one could both understand some emergent and collective phenomena in social systems and to build computer systems (decision-support tools), which aid individuals to take real time decisions receiving information from a complex environment. The weakness of ABM (or MAS) is their intrinsic complexity that acts as a counterpart of the reductionist approach of the hard sciences to model the physical reality. From one hand, this usually implies the presence of many parameters that have to be empirically evaluated taking advantage from the actual possibility of collecting big databases. From the other hand, it is impossible to take into account all the possible aspects of a complex socio-technological problem without reproducing the whole complexity of the reality. Consequently, the presence of over-fitting problems in the validation procedures of ABM is still a real risk. Under this point of view, the ABMs could be able to manage in an optimal way the empirical examples used in the validation process, but they could fail their goal when new unpredicted scenarios are proposed. This could be the case when congestion phenomena emerge in air traffic (or in general in a transportation system) or one has to face the effect of extreme events due to the lack of previous empirical observations. The investigation of the existence of physical laws (i.e. universal behaviors) that allow deeper understanding and explanation of the observed phenomena, is, on my opinion, still a key point in the*

future of Complex Systems Science. Even if the scientific method requires a reductionist approach, that it is not always justified in the case of complex systems, nevertheless it could point out the 'true' control parameters and the relevant observables to predict criticalities. Furthermore, it could suggest strategies to perform a control theory approach or to optimize in a robust and generic way the system performance in the framework of ABMs. In the specific problem of ATM, the predictability and the efficiency of the proposed ABM can be significantly improved by the results of a complex physics approach. ABMs have taken advantage from the continuously increasing of the computer performances to cope with large complex systems as the transportation systems. Nowadays they are one of the most promising approach to simulate complex socio-technological systems to understand the emergent properties of the considered systems related to the cognitive behavior of its components. In the definition and application of an ABM one has to face the following problems:

1. identify the all different agents that have to be modeled to simulate the considered system;
2. define the algorithmic procedures at the base of the agent behavioral strategies;
3. define the interaction between the cognitive behavior of the agents and the physical dynamics of the system;
4. perform the validation process using suitable empirical observations;
5. analyze the predictability of the ABM when new scenarios are presented.

*Only when the previous items are exhaustively studied, an ABM is a powerful tool ready to assist stakeholders to manage complex systems."*

Agent Based Models (ABMs) are therefore research tools used in several fields to observe, model and foresee scenario simulations of highly interconnected complex systems. Within the field of Air Traffic Management (ATM) research they are used to model present state of very complex socio-technical systems or sub-systems and to provide computer simulated scenarios of future settings. The final aim is to reach maturity in the ability of describing emergent phenomena in complex systems and use them as a decision-support tools for policy choices concerning the evolution and innovations to be introduced in the air transportation system.

**i) To what extent agent based models and multi-agents systems have been implemented in the study of the air traffic management complex system? Similarly, to what extent has existing research been able to demonstrate the value of these tools and what actions are required to move further towards industry adoption of such tools?**

One of the two experts has pointed out that the ELSA project has developed agent-based model of air traffic management system operations on a strategic and tactical level. Agents introduced at the strategic level are Airline Operators (AO) and Network Manager (NM) while at the tactical level agents are Aircraft/Pilots and Air Traffic Controllers (ATCo's). In the report it is stressed that a two-level model (strategic vs tactical) represents a reasonable approach because it describes a current way of working and because the outputs from strategic model (agreed flight plans between airlines and Network Manager) should be used as inputs into the tactical layer (actual realisation of flight plans). The ELSA model is of modular structure (contains few modules: flight list, conflict detection, conflict resolution, directs,

shocks, multi-sector) enabling decision makers (of various kinds) to perform experiments in different areas of interest.

According to the external experts the ELSA project has dealt with the first three points mentioned above, whereas further research work is necessary to accomplish the last two tasks. Moreover, it seems that from both strategic and tactical point of view no more agents are necessary, i.e. adequate agents are chosen to represent operations. Although experiments at the strategic level provide some interesting results, one of the experts highlights some issues related to the interactions amongst agents.

In the view of the external experts, the CASSIOPEIA project is a MAS, whose goal is to reproduce in silico the cognitive behavior of agents that rule the ATM providing both an interface with dynamical databases and the possibility to visualize and to interpret the simulation results. In the present state, the CASSIOPEIA MAS has mainly considered the aircraft agent, the airline agent and the airport agent to perform virtual reality experiments, which simulate different scenarios in the airline flight planning that could be relevant in the future when the air traffic demand is expected to increase. These scenarios are not directly connected to criticalities or congested states of transportation systems, but aim to find an optimized management of air traffic that minimizes the environmental impact, the economic cost of the delays. In order to enable wider usability of simulation platform, agents are defined on three levels: general (highest level), domain-specific and case-specific (lowest level). This approach is very beneficial because it allows a broad range of applications. Usability of software platform has been demonstrated through three different case studies, each of which has a different scope and agents involved.

The experts suggest that the model shows the possibility of simulating the very complex

interaction processes, which presides over the air traffic at departures and arrival at the airports. The cores of the model are aircraft agents and the airline agents that exchange continuously information with airports and the agents, which manage the arrivals and departures. One of experts suggests the introduction of new types of agents: the Network Manager (NM) and Air Navigation Service Providers (ANSP).

In the case of the work done by B. Monechi, the external experts noticed that ABM has been implemented at the level of Air Traffic Controllers (ATCo's) and conflict resolution (CR) behavior only. Others agents are not taken into account explicitly, e.g. airlines or pilots behavior during CR or their acceptance of trajectory changes. Related to ATCo's behavior during CR only horizontal resolution strategies are analyzed and modeled although vertical CR are more efficient and usually applied by ATCo's in current day operations. As a result, one of the experts highlights that other agents should also be included in the model as well as traffic flow management strategies applied by ATCo's and coordination procedure between ATCo's of adjacent sectors (which is simple modeled).

The main results of the Monechi's project consists in showing the existence of a generic scaling law, which relates the average flow on the air traffic network and the number of unsolved conflicts when the flow overcomes a certain threshold (the congestion threshold). The key point is that the exponent which characterizes the scaling law seems to be universal (i.e. independent from the details of the air traffic network considered), but both the congestion threshold and the exponent depend on the local strategies adopted by air controllers to solve the conflicts. These laws could not only constitute a useful tool to forecast and to control the congestion transitions in an air traffic network, but

also define utility functions in optimizing procedures of the network structure.

In conclusion, the experts' evaluation suggests that the considered projects have provided valuable contribution to the study of the ATM, and have successfully provided a sound proof of concept that ABMs are a fruitful tool to study this relevant socio-technical complex system. Each expert has focused his reports on specific aspects, so that further work can be pointed out to generalize these models in order to have a full picture of the system and in order to move the models from the proof of concept state to a state of an industrial product.

## **ii) What are the different actors of Air Traffic Management that can be most fruitfully modeled by ABMs and MASs?**

According to the external experts, the ELSA project might benefit from considering Airline Alliances at the strategic layer as well as Airports needs at the tactical level. Moreover, it might be beneficial to also consider the economic impact of the Network Manager choices on the competing companies.

Also for the CASSIOPEIA project, the experts suggest that the introduction of other agents should be beneficial. In this case one of them suggests the introduction of the Network Manager and Air Navigation Service Providers/ Air Traffic Controllers into the model.

In the work of B. Monechi the experts deem necessary to model airline/pilot behavior as well as airline trajectory negotiation with Network Manager and local ATC, as well as collecting data and modeling particular situations where the air traffic system has been close to a critical state due to failures in the air traffic network.

In conclusion, it seems that in general the choice (i) of the agents used to model the

ATM and (ii) of the mechanisms by which they interact has been done in a reasonable way, although more agents and interactions should be included in order to enhance the capability of the models to explain the ATM system.

## **iii) To what extent agent based models and multi-agents systems have shown a maturity level such to be used as part of decision-support tools MASs?**

According to the external experts, the ELSA project is of higher maturity and very close to industrial application as a decision-support tool. This model could be certainly a valid support to manage the air traffic in normal condition, reducing the workload of air-traffic controllers and the possible human error.

As for the CASSIOPEIA project, the experts suggest that the presented model provides good starting point for end users to build the models suitable for their problems. The platform is of middle to higher maturity and easily could be used as a decision-support tool, even if a validation process in specific cases is still necessary to quantify the predictability of the model.

The external experts instead think that the work of B. Monechi is not mature enough to be accepted by the industry, However one should consider that the developed model is part of a PhD thesis.

In conclusion, the two projects presented for the evaluation can be considered close to industrial applications.

## **iv) What are the aspects of the air traffic management system that are worthy of further investigation by making use of agent based Models and multi-agents systems?**

According to the external experts, the ELSA project might benefit from considering

- dynamic sectorisation of the airspace,

- the exploitation of different strategies for air traffic flow management,
- a better comparison with standard KPIs,
- the implementation of different economic strategies of the airlines at the strategic layer,
- the implementation of learning strategies especially at the tactical layer.

As for the CASSIOPEIA project, the experts suggest that the project might benefit from

- modeling the Network Manager behavior,
- modeling the ANSP/ATCO activities related to airlines and airports operations,
- a better comparison with standard KPIs,
- modeling of the passengers needs.

In the case of the B. Monechi work, the external experts suggest that it is necessary

- to model airline/pilot behavior during trajectory negotiation process at strategic level and trajectory change process at tactical level,
- modeling of coordination between ATCO's from adjacent sectors,
- modeling the Network Manager as additional agent interacting with local ATCOs.

Therefore, as mentioned above, there is room for introducing in the models more agents and interactions in order to enhance the capability of the models to explain the ATM system.

According to the external experts, the ELSA project is showing huge potential to be used as a decision-support tool. Its maturity is high which is bringing this model closer to industrial applications. Some further enhancements could be made by adding

some new agents in order to model a wider range of empirical facts. The algorithmic strategies at the roots of the ATM have been exhaustively analyzed in the ELSA project by considering many different real and artificial scenarios, even if some issues need further investigations.

As for the CASSIOPEIA project, the external experts think that developed ABM platform has huge potential to be used as a decision-support tool provided that further improvements are implemented, especially with the inclusion of new agents and the exploitation of solutions which better take into account the heterogeneity of the different actors involved.

According to the external experts, the work of B. Monechi is showing potential to be used as a decision-support tool provided that further improvements are implemented. Those improvements should see the inclusion of new agents as well as some operational procedures currently used or foreseen in future SESAR operational scenario. Moreover, the improvements should also regard the understanding of the physical aspects of congestion transition in transportation systems.

In conclusion, the judgment of the external experts on the way the application of Agent Based Models to the understanding of the ATM socio-technical complex system has been pursued within the Complex World project seems to be positive. Further work is still needed for a better understanding of many stylized facts in ATM, however, and more importantly, ABM seem to have the capability of addressing the open questions. In particular, it would be beneficial a validation phase in which the developed ABMs/MASs support the decisions of the controllers to see to which extent they prove their capability to manage efficiently a large ensemble of realistic scenarios.



## 5.F Integrating uncertainty into decision-support tools

There are many scenarios in ATM where uncertainty plays an important role. Examples of these include scheduling of arrivals/departures, routing around adverse weather, trajectory prediction, conflict resolution, and flow management. In the past, most integrated decision-support tools (DST) that have been developed to help manage these scenarios commonly neglect uncertainty. However, including the effect of uncertainty in DSTs might help to improve their efficiency, thus benefiting the ATM system.

There is not an unique way to include uncertainty in a DST. Possibilities range from considering the worst case scenario, buffers (such as intervals or confidence ellipsoids), Monte Carlo simulations, or more detailed statistical models.

However, there are many challenges in including uncertainty in a DST. For instance, it is not clear what type of statistical models should be used to realistically capture uncertainty. There is also a trade-off between robustness and performance: if one tries to accommodate too high levels of uncertainty, it might lead to excessive conservativeness in DST solutions. In addition, while in a deterministic setting an optimal solution is easy to define, this notion is not totally clear in an uncertain environment.

**i) Are the different sources of uncertainty in ATM properly identified and characterized? Is there any important uncertainty source that has been left out?**

It is clear that different scenarios would require different modelling of uncertainty sources. The research under review

investigates widely different examples with significantly varying uncertainty modelling. For instance, the ATFM is formulated including uncertainty in the future capacity of route and airport, which are assumed to be given in terms of static probabilities associated with capacity. However it is unclear how realistic this representation of uncertainty is or how the prescribed probabilities in capacity are determined in practice. The arrival scheduling problem is formulated as an optimisation problem on a graph, where the key parameters are the arrival and take off times; the authors consider uncertainty in these variables. The trajectory predictor computes uncertainty in estimated arrival time (ETA) due to uncertainties in the wind forecast, initial altitude errors, and errors due to the flight navigation system.

All these examples leave some important sources out of the picture, without proper justification in some cases. In route-planning problems, weather is an important factor which is oftentimes left out. Other relevant sources of uncertainty which are difficult to model and frequently neglected include aircraft malfunction/servicing, pilot/crew availability, departure/arrival delays, airborne holding, etc. TPs frequently do not consider uncertainty due to passenger boarding, gate delays, airport congestions which may significantly affect predicted ETA. On the other hand results from the literature show that in the context of runway optimisation, factors such weather, en-route flight delays, propagated delays from other airports, maintenance delays, etc., all manifest as uncertain delays in take off and arrival and do not need to be included explicitly.

## **ii) Have the researchers been able to demonstrate the value of including uncertainty in ATM models?**

The research under review shows that DSTs can include uncertainty in their models, but it is unclear in some cases how they lead to a more robust ATM. In the case of the ATFM algorithm, one reviewer points out that the resulting risk-adjusted plans are in fact more risky and less robust.

In the arrival scheduling problem, researchers show that deterministic DSTs might produce solutions that become infeasible in the presence of disturbances. Then they show that solving an optimal problem with uncertainty has much lower infeasibilities, thus proving the value of uncertainty inclusion. However, the stability gains do not appear very significant for the magnitude of the random departure delay considered and come at the expense of building additional delay into the plan.

## **iii) Is the description of uncertainty used by researchers adequate? Or is it too simple to be of significant value or too complex to be used in practice?**

According to the reviewers, the research under evaluation does not always use an adequate description of uncertainty. In the ATFM case, the reviewers point out that the representation is too simple, and difficult to compute from real data. This indicates that the researchers have not been able to properly justify their choice of representation. The arrival scheduling algorithm, on the other hand, proposes a sound statistical departure delay model which is validated against real delay data from a large German airport. The trajectory prediction problem is simplified and not validated, and the reviewers are concerned about the validity of the used techniques.

## **iv) To what extent the algorithms designed in these research projects would be**

## **implementable as DSTs? Would they improve current operations?**

The reviewers agree that the algorithms described in the documents could potentially be implementable. However, it is unclear if any of them would improve current operations. The ATFM algorithm may generate risky plans, and further research would need to be conducted to understand the benefits to the proposed approach in real operations.

The arrival scheduling algorithm could potentially be incorporated into an arrival manager (AMAN) to produce more robust arrival schedules that can contribute to more efficient utilisation of runways. The computational times are small enough (ten seconds), therefore it could be implemented in “real time”, where the planning is done over a window of ten minutes. It appears however that the effectiveness of the algorithms would depend on the quality of the delay statistics available for the specific airport where the algorithms are used.

The trajectory predictor uses Gaussian error propagation which is computationally simpler than a full scale Monte Carlo, especially for hybrid systems. Therefore, the proposed method would provide a computationally efficient framework. The predicted trajectory could be used to increase the robustness of decisions based on predicted trajectories, e.g. the detection of predicted conflicts. Besides, the researcher’s idea of measuring the allowable control authority of an FMS to compensate for a given time error at a point downstream the trajectory could be exploited to design new, more robust RTA algorithms. Unfortunately, it is unclear if the proposed method is in fact accurate for the problem considered here and it should be validated before it can be considered for practical application.

## v) Are the methods used to simulate scenarios with uncertainty scientifically sound?

The examples presented are academic, mainly designed to illustrate the proposed approaches. The simulation process is in general scientifically sound, based on Monte Carlo simulations with many scenarios. For instance, in the ATFM algorithm is tested in 1000 scenarios to verify constraint violations. From these runs, the authors conclude that risk-adjusted optimisation results far fewer infeasibilities than the problem with hard constraints. However, in this case it is difficult to interpret this conclusion as it is unclear how many of the 1000 scenarios are realistic or how the increased feasibility of the routing plan results in more robust ATM.

Similarly, the arrival scheduling algorithm uses Monte-Carlo techniques for validating the optimized assignments. It is unclear however how the assignments were validated. The trajectory predictor assumes in its Monte Carlo simulations re-defined probability distributions for the wind and initial altitude errors. Modelling errors are not considered as the reference “truth” trajectory is calculated with the same model assuming no wind or altitude uncertainty and therefore the applicability of the results in practice is limited.

Do researchers address the question of how to get a desired trade-off between robustness with respect uncertainty and performance degradation due to excessive conservativeness?

This was not addressed in most of the research under review and it is identified as a topic for further research. Only the arrival scheduling algorithm analyzes the trade-off between the increase in robustness of the arrival plan, i.e. number of required reassignments over the different planning cycles, achieved by considering

uncertainty explicitly, and the additional arrival delay that the algorithms built into the plan to achieve this additional stability. The researchers also consider other performance parameters such as computation time.

## Conclusions

The research under review shows that it is indeed possible to include uncertainty in DSTs, but that further research is still needed before they can be implemented in practice.

In particular, the reviewers believe that the RobustATM project is pioneering in considering delay uncertainty explicitly in the arrival scheduling problem and therefore it opens new avenues to design more robust algorithms for AMANs.

The results show that the approach works, for instance with one of the algorithms resulting in less reassignments but more delays than the other approaches.

The researchers acknowledge that the behaviour of the algorithms against larger delay disturbances needs to be investigated.

ONBOARD’s proposed approach to ATFM planning could potentially be implemented in a DST. However, in order to implement the proposed approach in practice, it would be necessary to be able to calculate the probability distributions of the sector capacities (DPDs), so that the chance constraints can be derived. The reviewers are concerned that the generation of risky plans actually results in a less robust ATM.

Alexander Schwithal’s research focused on trajectory prediction uncertainty from the point of view of the airborne trajectory, i.e. the trajectory predicted by the FMS. However, the findings of this research could be used in the context of ground-based TPs, since knowledge of the uncertainty of the airborne trajectory can be exploited on

the ground to increase the robustness of decisions based on predicted trajectories. The reviewers expressed concerns that the techniques used in this research might not be valid.

From the reviewer's input, some recommendations can be derived for further research in the topic:

- Uncertainty models used in DST might only consider some sources of uncertainty and neglect others, but these modelling choices have to be sound and based in the existing literature and real data.
- It is not sufficient to present a rigorous uncertain model, it is also important to be able to map this model to reality, this is, to have the capacity to extract the parameters of the model from real data.
- New algorithms have to convincingly demonstrate that they actually increase the robustness of ATM.
- Monte Carlo methods can be used to validate the algorithms, but realistic scenarios (extracted from real data) need to be used in this process. Otherwise the results are questionable.
- To assess the implementability of an algorithm, the trade-off between performance, robustness and computation time needs to be analyzed.

## 5.G Characterisation of meteorological uncertainty

It is well known the great impact of meteorological (MET) effects on ATM. These effects are ubiquitous for wind and temperature (nominal conditions); for instance, optimum routes for air traffic have a strong dependency on meteorological parameters such as the position of the jet stream or the strength and/or direction of prevailing winds. Moreover, in a very few cases and limited areas, MET hazards can potentially perturb the nominal traffic (significant weather conditions); indeed, adverse weather continues to be a major cause of delays in air travel.

However, accurate numerical weather prediction (NWP) forecast models continue to be challenging due to issues including uncertainty in observations used to initialise the forecasts and an incomplete understanding of the physical processes that occur in the atmosphere.

MET service providers have developed an innovative technique, referred to as Ensemble Weather Forecast (EWF), for identifying unpredictable weather episodes and precisely quantifying the related uncertainties. The basic concept is to run an ensemble of weather predictions while slightly altering the initial conditions and the parameters in the model that simulate unresolved physical processes. The ensemble, thus, constitutes a panel of possible evolutions of the weather situation. If weather is predictable all the solutions converge, whereas divergence is the signature of an unpredictable occurrence. Note, however, that the “most likely” forecast is not much improved with EWF. The added value is rather on the spread of the solutions in the ensemble. The 4D map of a MET parameter spread

allows end users to identify and track areas of high uncertainty, and the spread can be interpreted as an approximation of the parameter’s PDF at each grid point and time step of the model solutions. However, ensemble models tend to be underspread in that the actual uncertainty in the forecast is greater than what the ensemble members would suggest. This can be alleviated somewhat by constructing a ‘superensemble’ of ensemble models from several different modelling centres, or by using time-lagging to construct probabilities from several different runs of the same forecasting centre’s model.

Two main approaches can be considered when dealing with EWFs. First, the probabilistic approach, where EWFs can be analysed to calculate 4D maps of the probability distributions of parameters of interest (probabilistic MET forecast), that shall further be integrated in a probabilistic impact assessment model (provided that such a model exists). The probabilistic approach saves computer time. Spatial correlations, however, may be lost in maps of independent grid points probabilities, while MET uncertainties are mostly associated to mesoscale organised structures covering a large number of grid points at a scale larger than a control sector.

Alternatively, one has the ensemble approach, where each member of the MET forecast ensemble can be analysed by a deterministic impact assessment model. Then, a compact ensemble impact assessment can be built from which probability distributions can be derived. This approach leads to a large volume of data, especially when the number of

members is large, as one expects in the case of ‘superensembles’. Some type of post-processing is then required.

It is clear that there is no universal approach for all air traffic operational concepts. Selection criteria include the spatial nature of the process, Lagrangian or Eulerian (spatially correlated or not), the acceptable complexity of the calculations including its latency, and the user utility function as expressed for instance by a cost/loss ratio.

**i) Do researchers consider current, state-of-the-art meteorological models? Are they using real data to feed these models? Is meteorological uncertainty correctly included? Are the uncertain/stochastic models too simple to be of significant value / too complex to be used in practice? Are the methods used to simulate scenarios with uncertainty scientifically sound?**

In general, when using ensembles, one must ensure that caution is used in the T+0--T+6 hour period, where may be underspread due to the time taken for the initial condition perturbations and stochastic physics to spin up.

The process of calculating “optimum routes” from the ensemble must also deal with adverse weather information, because otherwise it might not always be possible to fly the routes found. Moreover, because in reality airline routes differ significantly from day to day, when applying the weather data for one particular day, one should consider the actual route the airline would have flown that day.

In the analysis of thunderstorms, it is advised to review the extensive literature on the field of verification of spatial objects. Also, besides using deterministic nowcast systems, it is recommended the use of ensemble-based probabilistic nowcast systems,

now available as part of the state of the art.

In ATFM it is recommended the use of the ensemble approach (rather than the probabilistic approach), because it allows to consider more than one weather parameter, as it is required in ATFM, which is likely to be affected by more than one weather parameter (e.g. winds, temperatures, visibility, convection, etc.).

**ii) To what extent the algorithms designed in these research projects would be implementable in real life, for instance as Decision-Support Tools? Would they improve current operations?**

In principle, more work is needed to guarantee manageable implementation in real-world situations of algorithms using probabilistic weather forecasts.

A great benefit of the implementation will be to discriminate between nominal conditions (most frequent situation) during which the expected level of performance of the new ATM concepts will be applicable and degraded weather conditions during which mitigation actions and regulations shall be implemented. In terms of KPI, one can reasonably expect a drastic reduction of unjustified regulations and an improved prediction of where and when they shall be implemented. The benefit of integrating information about MET forecast uncertainty in the planning phase thus aims at better anticipating the precise location (in space and time) of the airspace where mitigation will have to be applied with a level of confidence high enough to avoid unforeseen weather changes and last minute reactive safety actions with significant KPIs impacts.

**iii) When performing optimisation, do researchers address the question of how to get a desired trade-off between robustness (with respect to meteorological uncertainty) and performance degradation due to excessive conservativeness?**



This last point is not yet fully addressed in the documents and shall be considered as the priority in R&D.

## Conclusions

The conditions are fulfilled in SESAR for consolidating the dialogue between ATM and MET service providers to move from a reactive to a pro-active management of MET impacts on ATM. Ensemble weather prediction provides an efficient way of identifying areas of high uncertainty and it does not require, in principle, any modification of existing (deterministic) impact assessment tools, but just running them for each possible predicted future. Because areas of high forecast uncertainty are rare, it would be advisable to restrict the use of ensemble impact assessment to those few cases. From this perspective, MET forecast uncertainty is to be handled like adverse weather. This is the reason why the Met DOD in WP11.02 has been revised; “adverse weather conditions” has been replaced by “significant weather conditions” and MET forecast uncertainty has been introduced, together with adverse weather, in this category.

Nonetheless, despite the above discussion, probabilistic approaches are lacking, and therefore more research to advance in the development of probabilistic approaches seems to be needed.

The temporal dimension of the MET/ATM process predictability remains a key issue. MET requirements shall be defined that are consistent with ATM predictability. In SESAR1, MET requirements have been defined separately for medium, short term planning and execution. The issue of the consistency between these various time horizons must be considered as a whole to clarify how MET requirements shall evolve with the horizon of the forecast, spatial and time granularity of the information, refresh rate and the scale of the resolved processes.

Finally, it is clear that the integration of the MET forecast uncertainty into ATM is not trivial as it is highly dependent on the application. Hence, there is a need to find the best approach for each application, taking into account the compromise between the intrinsic complexity of the process and the efficiency of the application.

## 5.H Model-based identification of emergent behaviours at the design stage, including comparison with reality

Established system design takes a conservative approach regarding emergent behaviour by trying to avoid it. However this may be counterproductive because for a complex socio-technical system, it is impossible to identify and learn understanding emergent behaviour at all frequencies without conducting adequate simulations. As long as not all emergent behaviour is identified and understood it is unknown which are positive and which are negative. Though once understood, there is the possibility to adopt or strengthen positive emergent behaviour and to avoid or mitigate negative emergent behaviour. This means there is great design value in timely identifying positive and negative emergent behaviours of future socio-technical designs at frequencies ranging from regular to extremely rare.

In support of a step change in future ATM design, challenge H concerns model-based identification of emergent behaviours from early design stage on, including comparison with reality. This evaluation report aims to describe for this challenge H, which relevant achievements have been made within relevant SESAR research [1-8], and what the follow-up research directions are.

**i) How well has SESAR funded research addressed the topic of timely identifying positive and negative emergent behaviour of future ATM designs at various frequencies, i.e. both the frequent nominal as well as the rare non-nominal behaviours?**

The scope of this question has two complementary dimensions:

1) the advanced complexity science methods available to evaluate future socio-technical ATM designs; and 2) their application to various points in the huge design space of possible future socio-technical ATM designs.

Regarding the latter dimension, SESAR funded research [1-8] has considered a few ATM designs only. This means that the huge future ATM design space largely remains open for future exploration on emergent behaviours.

Regarding the former dimension, the SESAR funded research projects have explored the following four approaches in identifying emergent behaviours in ATM designs:

- Agent-based modelling and simulation (ABMS) [1],[4],[6-7]
- Network based simulation [5]
- Evolutionary search [2-3]
- Serious gaming [8]

Overall these works show a high level of computational sophistication and a deep level of detail and system understanding by the researchers. This high level of sophistication has typically been realized through preceding complexity science research outside SESAR. Moreover, these works demonstrate that each of these advanced methods make it possible to identify emergent behaviours that could not

have been identified through established fast time and human-in-the-loop simulations.

These works also show that there is a large variety in the specific types of issues

addressed by each of the four methods. Each of these four advanced methods has specific qualities in identifying specific type of emergent behaviour; this is depicted in the Table below.

EMERGENT BEHAVIOUR IDENTIFICATION ASPECT	AGENT-BASED MODELLING & SIMULATION	NETWORK-BASED SIMULATION	EVOLUTIONARY SEARCH	SERIOUS GAMING
Socio-technical scope	Y	-	Y	Y
Environmental uncertainties	Y	Y	-	-
Positive emergent behaviour	Y	Y	-	Y
Negative emergent behaviour	Y	-	Y	Y
Nominal emergent behaviour	Y	Y	-	Y
Rare emergent behaviour	Y	-	Y	-

With the exception of the network-based simulation methods, all advanced methods have shown to be able to address the socio-technical scope of ATM.

Only two of the four advanced methods have shown to be able to explicitly address environmental uncertainties which make the system dynamics and behaviour much more variable.

Three of the four methods have shown their capability in identifying positive emergent behaviour. Three of the four methods have shown their capability in identifying negative emergent behaviour. Two (Agent-based MC simulation and Serious Gaming) have shown to be able to identify both positive and negative emergent behaviours.

Similarly, three of the four methods have shown to be able in identifying nominal emergent behaviour, and two have shown to be able in identifying rare emergent behaviour. Only one method (Agent-based MC simulation) is able to do both.

There also is a large variety in the kind of ATM design changes that have been addressed within the different projects. This matches very well with the typical aim

of most early SESAR research projects to learn about the capability of the specific methods in identifying emergent behaviour in future ATM designs. In view of the rich capabilities that have been demonstrated for each of these four advanced methods, and the large size of the rather unexplored future ATM design space, this also means that the further exploration of this large ATM design space asks for many future applications of these methods.

## ii) How well has SESAR funded research addressed the topic of strengthening complexity science research regarding the socio-technical dimension of ATM?

The socio-technical dimension of ATM applications has been addressed by three of the four approaches. At the same time it has been shown within the MAREA project [4] that there are significant remaining challenges in further improvement of modelling the socio-technical dimension of ATM.

## iii) How well has SESAR funded research addressed the topic on developing systematic methods in assessing differences in inputs and outputs of a model and reality of the socio-technical air transport system?

Sensitivity analysis over the large parameter space has been demonstrated to work well for two of the four approaches (agent-based modelling and simulation and network based simulation). For the larger models it has also been identified that there is need for further development of approaches in evaluating uncertainties and differences between model and reality.

#### **iv) How well has SESAR funded research addressed the topic of developing convergent methods on the acceleration of rare event simulation methods with application to air transportation?**

Two of the four approaches considered made use of elsewhere developed MC simulation acceleration methods; these are Genetic Algorithm (GA) based search optimisation in [3], and Periodic Boundary Condition (PBC) in agent-based Monte Carlo (MC) simulation [6,7]. Because the specific ATM design applications are so special and demanding, there is a clear need for dedicated research on the development of novel methods for the further acceleration of MC simulations.

### **Conclusions**

Using advanced methods in the early identification of emergent behaviour in ATM designs is such new there are many directions and opportunities for improvements, additions and extensions of the initial SESAR conducted research for challenge H.

A summary of the most important directions are:

First of all, the results obtained show that there is good reason to apply the approaches to many more points in the huge future ATM design space.

A complementary further development is that in future air transport complexity

science work it is advisable that simulation code is being made available to other researchers (which is a common practice in many other domains).

A third direction is to combine the results from different complexity science approaches, which can range from a sequential use to an integrated and simultaneous use of complementary approaches (e.g. to exploit agent-based simulation tools within serious gaming with different stakeholders).

To strengthen the context-dependent modelling of the human behaviour and performance in the ATM loop in combination with empirically observed human behaviour, e.g. through serious gaming and conducting expert interviews.

To further the development and application of advanced techniques in the acceleration of Monte Carlo simulation methods, e.g. through sequential Monte Carlo simulation, using Latin Hypercube Sampling in sensitivity analysis, and dedicated high performance computing languages/facilities (e.g. CUDA, distributed computing).

To strengthen the verification and validation approaches regarding the developed models as well as the identified emergent behaviours.

### **References**

1. Blom, H.A.P., Stroeve, S.H., Bosse, T., Modelling of potential hazards in agent-based safety risk analysis, 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM2013), June 10-13, 2013, Chicago, Illinois, USA.
2. Blom, H.A.P., Bakker, G.J., In search of positive emergent behaviour in trajectory based operations, Proc. 3rd SESAR Innovation Days, KTH, Stockholm,

26-28 November 2013, 8 pages.

**3.** Blom, H.A.P., Bakker, G.J.,  
Can ground-based separation  
accommodate very high en route traffic  
demand as well as advanced self-  
separation? Proc. AIAA Conf. ATIO 2015,  
Dallas, 22-26 June 2015, AIAA 2015-3180,  
pp. 1-15.

**4.** Bouarfa, S., Blom, H.A.P., Curran, R.,  
Everdij, M.H.C., Agent-based modeling  
and simulation of emergent behavior  
in air transportation, Complex Adaptive  
Systems Modelling, Springer,  
Vol. 1, 2013, 26 pages.

**5.** Clegg, K., Alexander, R., Method  
description Technical Report, ASHICS  
deliverable 2.2, University of York,  
September 2012.

**6.** Clegg, K., Alexander, R., Model Analysis  
Technical Report, ASHICS deliverable 3.2,  
University of York, May 2013.

**7.** Faber, E., Aalmoes, R., de Groot, T.,  
Hryniewicz, R., Nieuwenhuisen, D.,  
Serious gaming for change in air traffic  
management, Proc. 26th ISPIM Conference,  
Shaping the frontiers of innovation  
management, Budapest, Hungary,  
June 14-17, 2015.

**8.** Sanchez, M., NEWO Final report,  
NEWO deliverable D11, Isdefe,  
September 2013.







